Using multilevel modelling to assess case-mix adjusters in consumers experience surveys in health care

Damman, O.C.; Stubbe, J.H.; Hendriks, M.; Arah, O.A.; Spreeuwenberg, P.; Delnoij, D.M.J.; Groenewegen, P.P.

Published in:
Medical Care

Using Multilevel Modeling to Assess Case-Mix Adjusters in Consumer Experience Surveys in Health Care

Olga C. Damman, MSc, Janine H. Stubbe, PhD,* Michelle Hendriks, PhD,* Onyebuchi A. Arah, MD, PhD,†‡ Peter Spreeuwenberg, MSc,* Diana M. J. Delnoij, PhD,§ and Peter P. Groenewegen, PhD*¶*

Background: Ratings on the quality of healthcare from the consumer's perspective need to be adjusted for consumer characteristics to ensure fair and accurate comparisons between healthcare providers or health plans. Although multilevel analysis is already considered an appropriate method for analyzing healthcare performance data, it has rarely been used to assess case-mix adjustment of such data. The purpose of this article is to investigate whether multilevel regression analysis is a useful tool to detect case-mix adjusters in consumer assessment of healthcare.

Methods: We used data on 11,539 consumers from 27 Dutch health plans, which were collected using the Dutch Consumer Quality Index health plan instrument. We conducted multilevel regression analyses of consumers' responses nested within health plans to assess the effects of consumer characteristics on consumer experience. We compared our findings to the results of another methodology: the impact factor approach, which combines the predictive effect of each case-mix variable with its heterogeneity across health plans.

Results: Both multilevel regression and impact factor analyses showed that age and education were the most important case-mix adjusters for consumer experience and ratings of health plans. With the exception of age, case-mix adjustment had little impact on the ranking of health plans.

Conclusions: On both theoretical and practical grounds, multilevel modeling is useful for adequate case-mix adjustment and analysis of performance ratings.

Key Words: consumer experiences, healthcare, case-mix adjustment, multilevel analysis, health plans

(*Med Care* 2009;47: 496–503)

From the *Netherlands Institute for Health Services Research, Utrecht, The Netherlands; †Department of Social Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; ‡Department of Epidemiology, University of California, Los Angeles (UCLA) School of Public Health, Los Angeles, California; §Centre for Consumer Experience in Health Care, Utrecht, The Netherlands; and ¶Department of Human Geography, Department of Sociology, Utrecht University, Utrecht, The Netherlands.

Financial support from the Dutch Ministry of Health, Welfare and Sport.

Reprints: Olga C. Damman, MSc, NIVEL (Netherlands Institute for Health Services Research), 3500 BN, Utrecht, The Netherlands. E-mail: o.damman@nivel.nl.

ISSN: 0025-7079/09/4704-0496

Public reporting of comparative healthcare information has become an important quality-improvement instrument in many countries.[1–3] In the Netherlands, Consumer Quality Index (CQ-index or CQI) instruments have been developed to assess quality of healthcare from the consumer's perspective.[4–7] CQI instruments are based on the American CAHPS (Consumer Assessment of Healthcare Providers and Systems) questionnaires[8–10] and Dutch QUOTE (QUality Of care Through the patient's Eyes) instruments,[11] which both measure consumers' experiences instead of inquiring after their satisfaction.

Theoretically, CQI ratings should be adjusted for a differential distribution of relevant consumer characteristics, so-called case-mix adjustment. This is important because, for example, some providers may receive poor ratings when they have many young consumers, who are generally less positive than older consumers.[9] Besides age, a variety of characteristics have been found to be associated with healthcare experiences: self-rated health status, education, sex, ethnicity, area of residence, income, language spoken at home, and health conditions.[4,9,12]

Different methods can be used to select consumer characteristics for adjusting experience scores and ratings. One method, used by CAHPS researchers in the United States,[9,12,13] combines heterogeneity (the distribution of a characteristic across providers) and predictive power (the amount of variance the characteristic predicts) into an impact factor. In research using CQI instruments, multilevel regression methods have been used to assess the performance of healthcare plans and providers and to investigate case-mix adjusters. This relatively new methodology has gained ground in analyzing hierarchical data in health services research.[14–18] Despite its known statistical properties for handling clustered data as often seen in consumer experience surveys and institutional performance assessment,[19–25] the multilevel regression method is rarely used for case-mix adjustment research.

The aim of this study is to investigate whether multilevel analysis is a useful tool to detect case-mix adjusters in consumer assessment of healthcare, and to compare this analysis with the impact factor approach.

Methods

Participants

Data collection took place in 2005 with the CQI health plan instrument "Experiences with Healthcare and Health

Insurer." For the current study, we performed secondary data-analyses of the experiences of 11,539 respondents from 27 health plans.

Measurement

The CQI health plan instrument consists of items on health plan services and received healthcare in the past year. It contains 54 core items on consumer experiences, 4 global ratings (family physician, specialist, healthcare, and health plan), 1 item on the likelihood to recommend plans to friends and family, and several items on consumer characteristics. The questionnaire is partly a transformation of the CAHPS 3.0 Adult Commercial Questionnaire.[5] We focused on 4 outcome variables (Table 1): the global rating of health plans and 3 experience scales (conduct of employees, health plan information, and reimbursement of claims) obtained from exploratory factor analysis of the experience items.

Statistical Analyses

The following 6 consumer characteristics were considered as case-mix adjusters: age, self-rated health status, education, sex, ethnicity, and urbanization of area of residence. We used 2 methodologies to explore which characteristics affect health plan experience domains and ratings: multilevel analysis and impact factor analysis.

Multilevel Regression Analysis

Multilevel linear regression analyses of consumers' experiences (N = 11,539) nested within health plans (N = 27) were performed. The first model contained no adjusters (model 0) and was the reference to which we compared other models with adjustments for only 1 consumer characteristic each (model 1 through model 6). A final seventh model adjusted for all characteristics. Both consumer and plan variance were estimated. We assessed the proportional changes in variance (PCV)[26] for the between-plan variance in each model to quantify the impact of adjustments on differences between plans. Specifically, the PCV was calculated as follows: absolute difference of the between-plan variance of the null model and the between-plan variance of the model with one or all characteristics included, divided by the between-plan variance of the null model. PCV's were calculated for plan variances only, because possible shifts in these variances reflect compositional or within-plan differences in the relevant consumer characteristic that influences plan ratings. A large PCV implicates that the characteristic is associated with relatively large alterations in the between-plan variance. In that case, quality rankings of plans are shifting, and the particular adjuster is relevant.

To illustrate the effects of adjustment on actual ratings, we considered the distribution of star ratings (*, worse than average, **, average and ***, better than average) for the global rating of health plans. This is a common method for presenting quality information, using plan means with comparison intervals,[27] and determining whether these intervals overlap with the overall mean across all health plans in the sample. We finally used Kendall's τ coefficients to measure the degree of correspondence between ordinal rankings of plans in different models.

Impact Factor Analysis

A consumer characteristic has impact when: (1) it is differentially distributed across health plan consumer populations (heterogeneity); and (2) it is associated with consumer experiences (predictive power).[12] The heterogeneity of each characteristic was calculated as the ratio of its between- and within-plan variance. Using traditional linear regressions, we estimated the predictive power of a specific consumer characteristic as the amount of variation predicted in a regression model including all consumer characteristics, minus the pre-

TABLE 1. Outcome Variables

Variable	No. Items	Items	Answering Categories	Cronbach's α
Global rating of health plan	1	Using any number from 0 to 10, where 0 is the worst health plan possible and 10 is the best health plan possible, what number would you use to rate your health plan?	0 to 10	
Conduct of employees	5	How often did your health plan's employees treat you with courtesy and respect? How often were your health plan's employees willing to help? How often did your health plan's employees listen carefully to you? How often did your health plan's employees explain things in a way that was easy to understand? How often did your health plan's employees spend enough time with you?	Never (1); Sometimes (2); Usually (3); Always (4)	$\alpha = 0.92$
Health plan information	3	How much of a problem was it for you to understand information that was mailed to you? How much of a problem was it for you to find information? How much of a problem was it for you to understand information that you found by yourself?	A big problem (1); A small problem (2); Not a problem (3)	$\alpha = 0.80$
Reimbursement of claims	2	How often did your health plan reimburse your claims in a short time period? How often did your health plan reimburse your claims correctly?	Never (1); Sometimes (2); Usually (3); Always (4)	$\alpha = 0.80$

dicted variation in a model excluding the specific characteristic. Dummies for health plans were included in both models. Predictive power and heterogeneity were multiplied and divided by a rescaling factor, correcting for differences in response scales of the various outcome variables. The number was also multiplied by 1000 for computational ease. As in previous research,[12,28] a case-mix adjuster with impact factor above 1 was considered important.

$$\textit{Impact factor} = (\textit{predictive power} \times \textit{heterogeneity} \times 1000)/\textit{rescaling factor}$$

For the same characteristic regressed on different outcome variables, a higher impact factor means that the characteristic has a higher effect on the outcome. For any 2 characteristics regressed on the same outcome, a difference in their impact factors implies a comparable difference in their effects on the outcome.

Results

Table 2 summarizes respondents' characteristics.

Multilevel Models

Table 3 describes the results of the multilevel regression analyses. The null model without adjustment showed significant variation between consumers and between health plans on all outcome variables.

For the global rating of health plans, the PCV's indicated that no more than 2% of the between-plan variances was explained by the included adjusters. The PCV for conduct of employees was 10% in all models including one characteristic, and 30% in the fully adjusted model. Concerning health plan information, adjusting for age only and later for all characteristics influenced the between-plan variance (PCV = 20%). The PCV for reimbursement of claims was 7% each in the model including education, as well as in the full model. In short, age and education seemed the most important adjusters.

Effect of Adjustments on Health Plan Ratings

Table 4 shows the shifts in star ratings on global rating of health plans in different models, compared with model 0. Adjusting for age had an impact on the ranking of 6 health plans. Kendall's τ coefficients showed positive significant correspondence between each model and model 0, indicating that rankings in different models did not differ significantly.

Impact Factor Analyses

Table 5 shows the impact factors of all consumer characteristics. Age had an impact factor of 6.31 on global rating of health plans and 2.56 on conduct of employees, implying that the age effect on the former outcome is 2 1/2 times its effect on the latter. Education showed an impact factor of 2.05 on global rating of health plans. No other consumer characteristic showed an impact factor of at least 1 on any outcome variable. Again, age and education seemed most important as case-mix adjusters.

TABLE 2. Person Characteristics of the 11,539 Respondents

Variable	N	%
Age, yrs		
18–24	774	6.7
25–34	1606	13.9
35–44	2327	20.2
45–54	2552	22.1
55–64	2330	20.2
65–74	1243	10.8
75 or older	707	6.1
Self-rated overall health status		
Excellent	1767	15.3
Very good	3034	26.3
Good	4791	41.5
Fair	1742	15.1
Poor	205	1.8
Sex		
Female	5717	49.5
Male	5822	50.5
Educational level		
1 (low: no primary education)	78	0.7
2	653	5.7
3	1910	16.6
4	404	3.5
5	1492	12.9
6	2273	19.7
7	1142	9.9
8	2587	22.4
9	812	7.0
10 (High: academic education)	188	1.6
Urbanization level		
1 (rural)	1902	16.5
2	2665	23.1
3	2244	19.4
4	2425	21.0
5 (big cities)	2303	20.0
Ethnicity		
Non-Dutch	689	6.0
Dutch	10850	94.0

Discussion

This study aimed to investigate the usefulness of multilevel regression for detecting case-mix adjusters of consumer experience data, in comparison to the commonly used impact factor analysis. Both multilevel regression and impact factor analyses of consumer experiences with Dutch health plans showed that age and education were the most relevant adjusters. Overall, case-mix adjustment did not have substantial impact on the ranking of most health plans and the distribution of star ratings. Nonetheless, using both statistical methods, it was shown that age and education slightly affected differences between health plans.

Although in this study both methods yielded similar results, the multilevel regression approach should be preferred on several statistical and practical grounds. First, given the hierarchical structure of consumer assessment data and the aim of

TABLE 3. Multilevel Analyses of Consumer Experiences With Health Plans

	Model 0 (Nul)	Model 1 (Age)	Model 2 (Health)	Model 3 (Education)	Model 4 (Sex)	Model 5 (Ethnicity)	Model 6 (Urbanization)	Model 7 (Full)
Global rating health plan (N = 11,276)								
Intercept	7.556 (0.060)*	7.401 (0.065)*	7.553 (0.063)*	7.420 (0.065)*	7.569 (0.061)*	7.567 (0.060)*	7.640 (0.065)*	7.522 (0.075)*
Age 18–24 (reference = 35–44)		−0.454 (0.053)*						−0.459 (0.054)*
Age 25–34		−0.189 (0.041)*						−0.158 (0.041)*
Age 45–54		0.073 (0.037)						0.082 (0.036)*
Age 55–64		0.305 (0.038)*						0.297 (0.038)*
Age 65–74		0.627 (0.045)*						0.613 (0.047)*
Age >75		1.013 (0.056)*						1.027 (0.060)*
Health status excellent (ref = very good)			0.000 (0.039)					0.065 (0.038)
Health status good			−0.013 (0.031)					−0.159 (0.030)*
Health status fair			0.062 (0.040)					−0.258 (0.040)*
Health status poor			−0.123 (0.095)					−0.455 (0.092)*
Education 1 (reference = education 6)				0.513 (0.152)*				0.479 (0.148)*
Education 2				0.723 (0.058)*				0.303 (0.060)*
Education 3				0.376 (0.040)*				0.201 (0.041)*
Education 4				0.020 (0.070)				0.013 (0.068)
Education 5				0.292 (0.043)*				0.097 (0.044)*
Education 7				−0.027 (0.047)				−0.011 (0.046)
Education 8				−0.009 (0.038)				−0.060 (0.037)
Education 9				−0.126 (0.054)*				−0.118 (0.053)*
Education 10				−0.199 (0.100)*				−0.262 (0.098)*
Sex (reference = male)					−0.028 (0.025)			0.017 (0.024)
Ethnicity (reference = Dutch)						−0.177 (0.052)*		−0.086 (0.051)
Urbanization 1 (ref = urbanization 5)							−0.235 (0.043)*	−0.168 (0.042)*
Urbanization 2							−0.100 (0.038)*	−0.089 (0.037)*
Urbanization 3							−0.063 (0.040)	−0.043 (0.038)
Urbanization 4							−0.032 (0.038)	−0.028 (0.037)
Variance consumers	1.691 (0.023)*	1.577 (0.021)*	1.690 (0.023)*	1.643 (0.022)*	1.690 (0.023)*	1.689 (0.023)*	1.685 (0.022)*	1.549 (0.021)*
Variance health plans	0.092 (0.026)*	0.094 (0.027)*	0.092 (0.026)*	0.092 (0.026)*	0.092 (0.026)*	0.091 (0.026)*	0.090 (0.026)*	0.090 (0.026)*
PCV†		2.17%	0.00%	0.00%	0.00%	1.09%	2.17%	2.17%
−2 log likelihood	38,003.350	37,220.470	37,997.360	37,683.250	38,002.080	37,991.930	37,968.790	37,021.720
$\chi^{2\ddagger}$		782.88*	5.99	320.10*	1.27	11.42*	34.56*	981.63*
Δdf		6	4	9	1	1	4	25
ICC§	0.052	0.056	0.052	0.053	0.052	0.051	0.051	0.055
Conduct of employees (N = 3793)								
Intercept	3.469 (0.022)*	3.445 (0.028)*	3.480 (0.027)*	3.411 (0.029)*	3.492 (0.024)*	3.486 (0.021)*	3.511 (0.030)*	3.473 (0.042)*
Age 18–24 (reference = 35–44)		−0.133 (0.042)*						−0.145 (0.042)*
Age 25–34		−0.058 (0.032)						−0.058 (0.032)
Age 45–54		0.032 (0.029)						0.052 (0.029)
Age 55–64		0.067 (0.032)*						0.096 (0.032)*
Age 65–74		0.138 (0.038)*						0.181 (0.039)*
Age >75		0.200 (0.050)*						0.287 (0.053)*
Health status excellent (ref = very good)			0.060 (0.033)					0.078 (0.032)*
Health status good			−0.006 (0.025)					−0.033 (0.025)
Health status fair			−0.057 (0.031)					−0.122 (0.032)*
Health status poor			−0.220 (0.027)*					−0.279 (0.064)*
Education 1 (reference = education 6)				−0.148 (0.137)				−0.076 (0.135)
Education 2				0.064 (0.059)				−0.011 (0.060)
Education 3				0.080 (0.035)*				0.037 (0.035)
Education 4				−0.049 (0.060)				−0.033 (0.059)
Education 5				0.105 (0.036)*				0.056 (0.037)

(*Continued*)

TABLE 3. (*Continued*)

	Model 0 (Nul)	Model 1 (Age)	Model 2 (Health)	Model 3 (Education)	Model 4 (Sex)	Model 5 (Ethnicity)	Model 6 (Urbanization)	Model 7 (Full)
Education 7				0.078 (0.036)*				0.090 (0.036)*
Education 8				0.081 (0.029)*				0.072 (0.029)*
Education 9				0.057 (0.041)				0.076 (0.041)
Education 10				0.072 (0.076)				0.064 (0.075)
Sex (reference = male)					−0.043 (0.020)*			−0.017 (0.020)
Ethnicity (reference = Dutch)						−0.234 (0.040)*		−0.212 (0.040)*
Urbanization 1 (ref = urbanization 5)							−0.094 (0.034)*	−0.060 (0.034)
Urbanization 2							−0.030 (0.031)	−0.026 (0.031)
Urbanization 3							−0.041 (0.032)	−0.032 (0.032)
Urbanization 4							−0.044 (0.032)	−0.040 (0.042)
Variance consumers	0.365 (0.008)*	0.360 (0.008)*	0.363 (0.008)*	0.364 (0.008)*	0.365 (0.008)*	0.362 (0.008)*	0.365 (0.008)*	0.350 (0.008)*
Variance health plans	0.010 (0.003)*	0.009 (0.003)*	0.009 (0.003)*	0.009 (0.003)*	0.009 (0.003)*	0.009 (0.003)*	0.009 (0.003)*	0.007 (0.003)*
PCV†		10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	30.00%
−2 log likelihood	6985.424	6924.409	6962.335	6967.140	6980.832	6950.717	6977.351	6821.702
χ^2		61.02*	23.09*	18.28*	4.59*	34.71*	8.07	163.72*
Δdf		6	4	9	1	1	4	25
ICC‡>	0.027	0.024	0.024	0.024	0.024	0.024	0.024	0.020
Health plan information (N = 2468)								
Intercept	2.613 (0.018)*	2.612 (0.024)*	2.626 (0.023)*	2.580 (0.026)*	2.610 (0.021)*	2.619 (0.018)*	2.615 (0.029)*	2.588 (0.042)*
Age 18–24 (reference = 35–44)		−0.106 (0.039)*						−0.127 (0.040)*
Age 25–34		0.021 (0.030)						0.022 (0.030)
Age 45–54		0.008 (0.029)						0.013 (0.029)
Age 55–64		0.037 (0.033)						0.048 (0.034)
Age 65–74		0.017 (0.050)						0.028 (0.051)
Age >75		−0.182 (0.107)						−0.143 (0.108)
Health status excellent (ref = very good)			0.051 (0.031)					0.057 (0.031)
Health status good			−0.025 (0.024)					−0.030 (0.025)
Health status fair			−0.063 (0.033)					−0.071 (0.034)*
Health status poor			−0.141 (0.088)					−0.107 (0.089)
Education 1 (ref = education 6)				−0.236 (0.227)				−0.148 (0.227)
Education 2				0.119 (0.107)				0.120 (0.108)
Education 3				0.006 (0.040)				0.001 (0.040)
Education 4				−0.006 (0.069)				0.010 (0.069)
Education 5				0.078 (0.042)				0.067 (0.042)
Education 7				0.066 (0.035)				0.093 (0.036)*
Education 8				0.042 (0.028)				0.033 (0.028)
Education 9				0.049 (0.039)				0.050 (0.040)
Education 10				−0.015 (0.066)				−0.021 (0.066)
Sex (reference = male)					0.007 (0.021)			0.012 (0.021)
Ethnicity (reference = Dutch)						−0.109 (0.047)*		−0.087 (0.047)
Urbanization 1 (ref = urbanization 5)							−0.048 (0.035)	−0.042 (0.036)
Urbanization 2							−0.013 (0.033)	−0.008 (0.033)
Urbanization 3							0.029 (0.034)	0.029 (0.034)
Urbanization 4							0.020 (0.034)	0.021 (0.033)
Variance consumers	0.253 (0.007)*	0.251 (0.007)*	0.251 (0.007)*	0.252 (0.007)*	0.253 (0.007)*	0.252 (0.007)*	0.252 (0.007)*	0.247 (0.007)*
Variance health plans	0.005 (0.002)*	0.004 (0.002)*	0.005 (0.002)*	0.005 (0.002)*	0.005 (0.002)*	0.005 (0.002)*	0.005 (0.002)*	0.004 (0.002)*
PCV†		20.00%	0.00%	0.00%	0.00%	0.00%	0.00%	20.00%
−2 log likelihood	3633.824	3617.291	3620.935	3624.323	3633.720	3628.479	3627.020	3580.816
$\chi^{2\ddagger}$		16.53*	12.89*	9.50	0.10	5.35*	6.80	53.01*
Δdf		6	4	9	1	1	4	25
ICC§	0.019	0.016	0.020	0.019	0.019	0.019	0.019	0.016

(*Continued*)

TABLE 3. (*Continued*)

	Model 0 (Nul)	Model 1 (Age)	Model 2 (Health)	Model 3 (Education)	Model 4 (Sex)	Model 5 (Ethnicity)	Model 6 (Urbanization)	Model 7 (Full)
Reimbursement of claims (N = 7359)								
Intercept	3.663 (0.025)*	3.627 (0.027)*	3.668 (0.027)*	3.643 (0.028)*	3.690 (0.025)*	3.674 (0.025)*	3.692 (0.027)*	3.675 (0.032)*
Age 18–24 (ref = 35–44)		−0.109 (0.029)*						−0.111 (0.029)*
Age 25–34		−0.003 (0.020)						0.010 (0.020)
Age 45–54		0.034 (0.018)						0.039 (0.018)*
Age 55–64		0.080 (0.018)*						0.087 (0.019)*
Age 65–74		0.117 (0.023)*						0.127 (0.023)*
Age >75		0.145 (0.031)*						0.182 (0.032)*
Health status excellent (ref = very good)			0.046 (0.020)*					0.060 (0.019)*
Health status good			−0.005 (0.015)					−0.026 (0.015)
Health status fair			−0.037 (0.019)*					−0.077 (0.020)*
Health status poor			−0.171 (0.047)*					−0.211 (0.047)*
Education 1 (reference = education 6)				−0.177 (0.094)				−0.138 (0.093)
Education 2				0.096 (0.035)*				0.059 (0.036)
Education 3				0.073 (0.021)*				0.048 (0.021)*
Education 4				−0.026 (0.035)				−0.018 (0.034)
Education 5				0.044 (0.021)*				0.025 (0.022)
Education 7				0.018 (0.022)				0.036 (0.022)
Education 8				0.011 (0.018)				0.007 (0.018)
Education 9				−0.036 (0.026)				−0.024 (0.026)
Education 10				−0.054 (0.048)				−0.058 (0.048)
Sex (reference = male)					−0.054 (0.012)*			−0.042 (0.012)*
Ethnicity (ref = Dutch)						−0.195 (0.027)*		−0.174 (0.027)*
Urbanization 1 (ref = urbanization 5)							−0.087 (0.021)*	−0.056 (0.021)*
Urbanization 2							−0.030 (0.019)	−0.024 (0.018)
Urbanization 3							−0.029 (0.019)	−0.020 (0.019)
Urbanization 4							−0.008 (0.019)	−0.001 (0.018)
Variance consumers	0.261 (0.004)*	0.258 (0.004)*	0.260 (0.004)*	0.259 (0.004)*	0.260 (0.004)*	0.259 (0.004)*	0.260 (0.004)*	0.253 (0.004)*
Variance health plans	0.015 (0.005)*	0.015 (0.004)*	0.015 (0.004)*	0.016 (0.005)*	0.015 (0.004)*	0.015 (0.004)*	0.015 (0.004)*	0.014 (0.004)*
PCV†		0.00%	0.00%	6.67%	0.00%	0.00%	0.00%	6.67%
−2 log likelihood	11,061.850	10,976.020	11,034.490	11,025.940	11,042.180	11,010.020	11,042.290	10,828.090
$\chi^{2\ddagger}$		85.83*	27.36*	35.91*	19.67*	51.83*	19.56*	233.76*
Δdf		6	4	9	1	1	4	25
ICC§	0.054	0.055	0.055	0.058	0.055	0.055	0.055	0.052

The categories with the second most respondents were chosen as reference categories; regression coefficients were estimated (standard errors in parentheses); the number of respondents differed across the 4 outcome variables, because of the fact that certain questions in the questionnaire were not applicable to all consumers.

*$P < 0.05$.

†Proportional change in variance.

‡Changes in −2 log likelihood were calculated; significance testing with χ^2.

§Intraclass correlation = Variance health plans/(Variance health plans + Variance consumers).

institutional profiling, the multilevel regression approach is needed to handle the within-group clustering.[17,29–32] The traditional linear regression used in the impact factor analysis assumes independence of consumer experience observations. This leads to biased standard errors used in the hypothesis testing of the main effects of the potential case-mix adjusters. The use of dummies for the groups (health plans) does not solve the failure of the independence assumption here. This can lead to biased differences in provider ratings and effects of case-mix adjusters.[29–31]

Second, impact factor analysis becomes increasingly inefficient when large numbers of group units are involved, because it uses many dummy variables to adjust for group effects. In this study, 26 (= 27 − 1) health plan dummies were used in the impact factor analysis. In larger studies, such as a planned CQ-index measurement of more than 4000 family practices, using a large amount of provider dummies is undesirable as it leads to inefficiency and model instability.

Third, unlike in the traditional regression of impact factor analysis, multilevel regression estimates are "shrunken" toward the population mean and give more precise and accurate predictions.[32] This guards against extreme values from small numbers of cases within particular providers. Fourth, the multilevel regression method is less labor intensive than the impact factor analysis. Finally, multilevel analysis enables us to detect effects of adjustment on ratings themselves, as was done in this study for star ratings. After all, this is the information presented to consumers.

Concerning case-mix adjustment in general, we believe efforts should be made to ensure that performance scores

TABLE 4. Changes in Ranking of Health Plans in Different Multilevel Models Compared With a Model Without Any Adjustments (Null Model) and Kendall's τ Coefficients

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Star ratings global rating health plan							
1 remains 1	8	9	9	9	9	8	7
2 remains 2	6	11	9	11	11	10	6
3 remains 3	7	7	7	7	7	7	7
1 becomes 2	1	0	0	0	0	1	1
2 becomes 1	4	0	1	0	0	1	4
2 becomes 3	1	0	1	0	0	0	1
3 becomes 2	0	0	0	0	0	0	0
Kendall's τ coefficient							
Global rating health plan	0.81*	0.99*	0.81*	0.98*	0.98*	0.93*	0.81*
Conduct of employees	0.89*	0.97*	0.96*	0.98*	0.97*	0.96*	0.80*
Health plan information	0.90*	0.89*	0.94*	0.98*	0.93*	0.92*	0.81*
Reimbursement of claims	0.93*	0.98*	0.95*	1.00*	0.98*	0.96*	0.91*

*$P < 0.01$.

TABLE 5. Heterogeneity, Predictive Power, and Impact Factor of the 6 Consumer Characteristics for the 4 Outcome Variables*

	Age	Health	Education	Sex	Ethnicity	Urbanization
Heterogeneity†						
Global rating health plan	0.052	0.030	0.123	0.050	0.035	0.153
Conduct of employees	0.051	0.032	0.097	0.042	0.040	0.138
Health plan information	0.048	0.017	0.113	0.049	0.009	0.124
Reimbursement of claims	0.063	0.026	0.096	0.059	0.035	0.136
Predictive power						
Global rating health plan	0.051	0.007	0.007	0.000	0.000	0.001
Conduct of employees	0.023	0.011	0.002	0.001	0.008	0.001
Health plan information	0.002	0.005	0.000	0.000	0.001	0.000
Reimbursement of claims	0.011	0.006	0.001	0.001	0.005	0.001
Impact factor						
Global rating health plan (RF = 0.420)‡	6.314	0.500	2.050	0.000	0.000	0.364
Conduct of employees (RF = 0.458)‡	2.561	0.769	0.424	0.092	0.699	0.301
Health plan information (RF=0.700)‡	0.137	0.121	0.000	0.000	0.013	0.000
Reimbursement of claims (RF = 0.696)‡	0.996	0.224	0.138	0.085	0.251	0.195

*The core model assumptions such as linearity and distributions in the impact factor analysis are the same as for traditional linear regression models. The impact factor approach assumes that missing-data mechanism is missingness-at-random given available variables and that using health-plan dummies or so-called fixed effects effectively addresses health-plan variability.

†Both between plan and within plan variance were estimated for each characteristic in linear mixed models by "intercept variance" and "residual variance," respectively. The consumer characteristic of interest was the dependent variable and the data were permitted to have a correlated and non-constant covariance matrix.

‡The rescaling factor (RF) was calculated based on the variance of each outcome variable. The numerator of the RF is the variance of the aggregated mean on the outcome variable. The denominator of the RF is the variance of the unstandardized predicted value in a linear regression model with all consumer characteristics and dummies for health plans included on the same outcome variable.

reflect health plans' actual performance, and not compositional issues arising from their differential consumer profiles. Given a healthcare market in which healthcare plans and providers are held accountable for their performances, even seemingly small adjustments are important for fair comparisons. Although we had no information on other characteristics than the self-reported characteristics under consideration, we recognize that other factors, such as disease status and severity, comorbidities, and prior healthcare utilization, might be more predictive and should be investigated in future research.[12] For example, administrative claims data could be tested to assess the effect of expected use of healthcare. However, variables like healthcare utilization should not always be adjusted for, because health plans might influence utilization through regulating access to healthcare. The aim of case-mix adjustment in CQ-index measurements is not to explain differences between healthcare plans or providers, but to ensure fair comparisons. Statistical adjustment should therefore only be conducted after careful theoretical and policy considerations, and only for variables that plans or providers cannot influence themselves.[33,34]

If there is any suspicion that a case-mix adjuster also adjusts for systematic differences in the quality of services that different consumer groups receive, it is better to refrain from such adjustment, and to present unadjusted data for these groups separately instead or to search for other methods. In this respect, it may be worthwhile to explore the possibility of using anchoring vignettes for the "calibration" of responses as an alternative for case-mix adjustment.[35] Meanwhile, we argue that properties of the multilevel regression method make it an appropriate tool for examining both case-mix adjustment and performance analysis of consumer experience data, especially given the clustered, frequently unbalanced, and sometimes sparse nature of such data.

ACKNOWLEDGMENTS

The authors thank Professor Juan Merlo for his insightful comments on the MLA methodology.

REFERENCES

1. Marshall MN, Brook RH. Public reporting of comparative information about quality of healthcare. *Med J Aust*. 2002;176:205–206.
2. Marshall MN, Shekelle PG, Davies HT, et al. Public reporting on quality in the United States and the United Kingdom. *Health Aff*. 2003;22:134–148.
3. Mannion R, Goddard M. Public disclosure of comparative clinical performance data: lessons from the Scottish experience. *J Eval Clin Pract*. 2003;9:277–286.
4. Arah OA, ten Asbroek AH, Delnoij DM, et al. Psychometric properties of the Dutch version of the Hospital level Consumer Assessment of Health Plans Survey instrument. *Health Serv Res*. 2006;41:284–301.
5. Delnoij DM, ten Asbroek G, Arah OA, et al. Made in the USA: the import of American Consumer Assessment of Health Plan Surveys (CAHPS) into the Dutch social insurance system. *Eur J Public Health*. 2006;16:652–659.
6. Stubbe JH, Brouwer W, Delnoij DM. Patients' experiences with quality of hospital care: the Consumer Quality Index Cataract Questionnaire. *BMC Ophthalmol*. 2007;7:14.
7. Stubbe JH, Gelsema T, Delnoij DM. The Consumer Quality Index Hip Knee Questionnaire measuring patients' experiences with quality of care after a total hip or knee arthroplasty. *BMC Health Serv Res*. 2007;7:60.
8. Cleary PD, Edgman-Levitan S. Healthcare quality. Incorporating consumer perspectives. *JAMA*. 1997;278:1608–1612.
9. Zaslavsky AM, Zaborski LB, Ding L, et al. Adjusting performance measures to ensure equitable plan comparisons. *Healthcare Financ Rev*. 2001;22:109–126.
10. Zaslavsky AM, Zaborski LB, Cleary PD. Factors affecting response rates to the Consumer Assessment of Health Plans Study survey. *Med Care*. 2002;40:485–499.
11. Sixma HJ, Kerssens JJ, Campen CV, et al. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. *Health Expect*. 1998;1:82–95.
12. O'Malley AJ, Zaslavsky AM, Elliott MN, et al. Case-mix adjustment of the CAHPS Hospital Survey. *Health Serv Res*. 2005;40:2162–2181.
13. Kim M, Zaslavsky AM, Cleary PD. Adjusting Pediatric Consumer Assessment of Health Plans Study (CAHPS) Scores to Ensure Fair Comparison of Health Plan Performances. *Med Care*. 2005;43:44–52.
14. Leyland AH, Groenewegen PP. Multilevel modelling and public health policy. *Scand J Public Health*. 2003;31:267–274.
15. Arling G, Lewis T, Kane RL, et al. Improving quality assessment through multilevel modeling: the case of nursing home compare. *Health Serv Res*. 2007;42:1177–1199.
16. Frick U, Rehm J, Krischker S, et al. Length of stay in a German psychiatric hospital as a function of patient and organizational characteristics—a multilevel analysis. *Int J Methods Psychiatr Res*. 1999;8:146–161.
17. Zaslavsky AM. Using hierarchical models to attribute sources of variation in consumer assessments of healthcare. *Stat Med*. 2007;1900:26:1885.
18. Davis P, Lay-Yee R, Scott A, et al. Do hospital bed reduction and multiple system reform affect patient mortality?: a trend and multilevel analysis in New Zealand over the period 1988–2001. *Med Care*. 2007;45:1186–1194.
19. Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J Royal Stat Soc*. 1996;159:385–443.
20. Leyland AH, Boddy FA. Measuring performance in hospital care. Length of stay in gynaecology. *Eur J Public Health*. 1997;7:136–143.
21. Christiansen CL, Morris CN. Improving the statistical approach to healthcare provider profiling. *Ann Intern Med*. 1997;127:764–768.
22. Burgess JF Jr, Christiansen CL, Michalak SE, et al. Medical profiling: improving standards and risk adjustments using hierarchical models. *J Health Econ*. 2000;19:291.
23. Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health*. 2005;59:438–439.
24. Merlo J, Chaix B, Ohlsson H, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health*. 2006;60:290–297.
25. Merlo J. Changing analytical approaches in European epidemiology—a short comment on a recent article. *Eur J Epidemiol*. 2005;20:737.
26. Merlo J, Gerdtham UG, Eckerlund I, et al. Hospital level of care and neonatal mortality in low- and high-risk deliveries: reassessing the question in Sweden by multilevel analysis. *Med Care*. 2005;43:1092–1100.
27. Goldstein H, Healy MJR. The graphical presentation of a collection of means. *J Royal Stat Soc*. 1995;581:175–177.
28. Arah OA, Rupp I, Delnoij DMJ, et al. Case mix adjustment strategy for the Dutch version of the Hospital CAHPS Instrument. In: AcademyHealth 2006 Annual Research Meeting. Seattle, WA: Academy Health; 2006.
29. Normand ST, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. *Stat Sci*. 2007;22:206–226.
30. Daniels MJ, Gatsonis C. Hierarchical generalized linear models in the analysis of variance in healthcare utilization. *J Am Stat Assoc*. 1999;94:29–42.
31. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press; 2007: chapters 1 and 11.
32. Snijders TAB, Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. 1st ed. Thousand Oaks, CA: SAGE Publications; 1999.
33. Zaslavsky AM. Issues in case-mix adjustment of measures of the quality of health plans. In: Proceedings, Section on Social Statistics. Alexandria, VA: American Statistical Association, 1998:56:–64.
34. Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes*. 3rd ed. Chicago, IL: Health Administration Press; 2003.
35. King G, Murray CJL, Salomon JA, et al. Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am Polit Sci Rev*. 2003;97:567–584.